PNAS
www.pnas.org

Supplementary Information for

Rationally engineered *Staphylococcus aureus* Cas9 nucleases with high genome-wide specificity

Yuanyan Tan*, Athena H. Y. Chu*, Siyu Bao, Duc Anh Hoang, Firaol Tamiru Kebede, Wenjun Xiong, Mingfang Ji, Jiahai Shi, Zongli Zheng

Correspondence: Jiahai Shi and Zongli Zheng
Email: Jiahai.Shi@cityu.edu.hk; ZhengZongli@gmail.com

This PDF file includes:

- Supplementary Material and Methods
- Figures S1 to S8
- Tables S1 to S5
- References

www.pnas.org/cgi/doi/10.1073/pnas.1906843116

Supplementary Material and Methods

Protein structure analysis Pymol (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.) was used to view protein structure of SaCas9 (PDB ID 5AXW and 5CZZ) and calculate polar contact between proteins and DNA.

Plasmids and oligonucleotides. The plasmid BPK2139 (Addgene #65776) was used to express WT-SaCas9 and variant mutagenesis. All mutant SaCas9 (including S-HF) expression plasmids (Table S1) were constructed by site-directed mutagenesis using BPK2139 as the backbone (Guangzhou IGE biotechnology Ltd.). High-fidelity SpCas9 variant expressing plasmids were purchased from Addgene (eSpCas9(1.1), #71814, SpCas9-HF1 #72247 and HyPa-Cas9 #101178). Sequences of mutated plasmids were confirmed by Sanger sequencing. The sgRNA expression plasmids for SaCas9 were constructed by ligating oligonucleotide duplexes (synthesized by BGI Tech Solution Co. Ltd.) of the sgRNA (Table S2, S3) into BsmBI digested BPK2660 plasmid (Addgene #70709). The sgRNA expression plasmids for SpCas9 were constructed by ligating sgRNA into BsmBI digested BPK1520 plasmid (Addgene #65777). All plasmids were purified with the PureLinkTM HiPure Plasmid Midiprep Kit (Invitrogen) and quantified by NanoDrop 2000.

Cell culture and transfection. The 293T cells were purchased from ATCC (CRL-3216), and the 293T-EGFP cell was a gift from Dr. Minh Le of City University of Hong Kong. Cells were cultured in advanced DMEM medium (Life Technologies) with 10% fetal bovine serum (Biosera) and 1% antibiotic (Life Technologies) in 5% CO_2 cell culture incubator. Transfection were performed using PEI (***Polyethylenimine***) for 293T following the manufacturer's instruction. 7.5×10^4 cells were transfected with a total of 400 ng SaCas9 plasmid DNA and 400 ng sgRNA plasmid in 24-well plate. WT-SaCas9 was co-transfected with a U6-null BPK2660 plasmid as the no-sgRNA negative control.

DNA extraction. Genomic DNA were extracted using the MiniBEST Universal Genomic DNA Extraction Kit (TaKaRa), and quantified using the Qubit 3.0 fluorometer and Qubit™ dsDNA HS Assay Kit (Invitrogen).

Targeted deep sequencing. Genomic regions (listed in Table S2) encompassing on-target sites of the 27 sgRNA examined (Table S2, S3) and a selection of off-target sites targeted by VEGFA_8, FANCF_13, EMX1_6 and RUNX1_13 were subjected to amplicon sequencing to evaluate editing efficiency of the SaCas9. Specifically, genomic DNA (20 ng for on-target site and 10 ng for off-target site) was used for PCR using target specific forward/reversed primers (Table S4) and Platinum Taq DNA polymerase (Thermo Fisher Scientific) by the following condition: 95°C for 5 min, 14 cycles of [95°C for 30 sec, 72°C touchdown at -1°C per cycle for 1 min], 20 cycles of [95°C for 30 sec, 58°C for 1 min], 72°C for 3 min and hold at 4°C. The product was used for a second PCR step to introduce Illumina sequencing sequences by the following condition: 95°C for 5 min, 20 cycles of [95°C for 30 sec, 65°C for 1 min], 72°C for 3 min and hold at 4°C. The sequencing libraries were quantified by qPCR (KAPA Library Quantification Kits for Illumina) and sequenced on Illumina NextSeq 500 System using 300-cycle kit

for 2x150 sequencing. The FASTQ reads were aligned to the human reference genome GrCh37 using BWA MEM[1] with default parameters. High-quality mapped read pair (properly mapped read-pair that contain at least 30 bp of matching alignment to the reference genome) were used for assessing the editing efficiency. Editing efficiency for each SaCas9 variant was measured as the number of reads containing InDels within the surveyed site, excluding the PCR primer annealing regions, divided by the total number of reads spanning the surveyed site, using an in-house Python script.

EGFP disruption assay. EGFP disruption assays were performed to assess on-target activity of wild-type and mutant SaCas9s with the sgRNA targeted to EGPF gene in the 293T-EGFP cell. Briefly, 7.5×10^4 293T-EGFP cells were transfected with 500 ng WT or mutant SaCas9 expressing plasmid and 250 ng of EGFP targeted sgRNA (either with or without mismatches to the target site) expression plasmid. The 293T-EGFP cell co-transfected with SaCas9 expression plasmid and the U6-null plasmid was used as baseline negative control for SaCas9 editing. The fluorescence of the transfected cells was measured by flow cytometry. On-target efficiency is measured as the loss of fluorescence upon co-transfection of sgRNA and SaCas9 expression plasmids and normalized to the mentioned negative control.

GUIDE-seq. Genomic DNA were extracted 72 h post-transfection and 400 ng was used for NGS library construction following the GUIDE-seq[2] methods with minor modification. Briefly, DNA was enzymatically fragmented by KAPA Frag Kit (KAPA Biosystems), followed by adaptor ligation and two rounds of hemi-nested PCR enrichment for dsODN integration fragments. To unify Illumina sequencing workflows for obtaining dual indexed data using Single-Indexed sequencing workflow across various Illumina platforms, we redesigned the original half-functional adaptors[3] and placed sample index (Index 2) at the head of Read 1, following unique molecular index (Table S5). Final sequencing libraries were quantified by qPCR (KAPA Library Quantification Kits for Illumina) and sequenced on Illumina NextSeq 500 System. Data demultiplexing of Index 1 was performed by bcl2fq v2.19, followed by custom scripts for Index 2 demultiplexing, adaptor trimming using the BBduk tool[4], and formatting for analysis using the GUIDE-seq software[5]. Briefly, demultiplexed and unique molecular index (UMI)-tagged FASTQ data was consolidated to generate UMI-consensus sequence, and aligned to human reference genome GrCh37 using BWA MEM[1]. High-quality alignments (MAPQ $\geq$ 50) were used for identifying genomic regions with the dsODNs integration as SaCas9 edited sites. Off-target sites with up to 7 mismatches edited by SaCas9, and 6 mismatches edited by SpCas9 within the protospacer region were identified.

AAV transduction The plasmid pAAV-CMV-SaCas9-2A-mCherry-U6-BsaI-sgRNA was cloned by PCR amplifying the 2A-mCherry fragment (Forward primer: 5'-CGCGGATCCGAGGGCAGAGGCAG-3'; reverse primer: 5'-CCGGAATTCTTACTTGTACAGCTCGTCCATGC-3') and ligating it via BamHI/EcoRI sites into the backbone plasmid pAAV-CMV-SaCas9-U6-BsaI-sgRNA (Addgene #61591). The WT-SaCas9 human optimized codon (pCAG-WT-SaCas9) was obtained from Addgene (Addgene #65776). The WT-SaCas9 or SaCas9-HF was cloned

into the pAAV-CMV-SaCas9-2A-mCherry-U6-BsaI-sgRNA to replace the original SaCas9 via AgeI/XhoI sites. The *VEGFA* gRNA was inserted into the pAAV backbone via BsaI sites. The control vector pAAV-CMV-mCherry-U6 *Rho* gRNA was used. For packaging AAV8 virus, pAAV rep/Cap2/8 and adenovirus helper plasmids were obtained from the University of Pennsylvania Vector Core (Philadelphia, USA). Recombinant AAV8 vectors were produced as previously described[6]. Briefly, pAAV vector plasmid, rep/cap 2/8 packaging plasmid, and adenoviral helper plasmid were mixed with polyethylenimine (PEI) with a ratio of DNA:PEI=1:3 in DMEM. The mixture was incubated 15 min at room temperature and added to HEK293T cells (HCL4517; Thermo Scientific) cultured in DMEM with 2% NuSerum growth medium supplement (Corning). The transfected medium was replaced by DMEM after 24 hours, and the supernatant was collected at 72 hours after transfection. The AAV8 virus supernatant was precipitated (mixed with 8.5% wt/vol PEG-8000 and 0.4 M NaCl for 2h at 4 °C), centrifuged at 7,000 × *g* for 10 min, and resuspended in virus buffer (150 mM NaCl and 20 mM Tris, pH 8.0). The virus supernatant was ultra-centrifuged under iodixanol gradient at 147,000 x *g* at 4ºC for 90 min. The collected fraction of AAV8 vectors (40% iodixanol fraction) were washed three times with PBS using Amicon 100K columns (EMD Millipore). The virus titers were quantified by protein SDS-PAGE gel and SYPRO® Ruby protein gel stain (Thermo Fisher Scientific).

ARPE-19 (ATCC® CRL-2302™) cells were cultured in DMEM/F12 with 10%FBS and incubated at 37 ºC , 5% CO2. ARPE-19 cells were transduced with AAV8 vectors as described by the Viral Vector Core Facility, University of Iowa Health Care. In brief, the AAV8 vectors were diluted in transduction medium (DMEM/F12, 2% FBS, 2μM Hoechst-33342) at room temperature and inoculated at 10^5 viral genomes per cell (vg/cell) of ARPE-19 cells incubated at 37ºC, 5% CO_2. The cells incubated with the transduction medium and without AAV vector transduction were used as the untransduction control. The transduction medium was replaced with full medium (DMEM/F12, 10% FBS) at 24 hours after transduction. The cells were harvested for genomic DNA extraction at 72 hours after transduction. Genomic DNAs were extracted by PureLink™ Genomic DNA Mini Kit (Thermo Fisher Scientific).

Data analysis. Sequencing data is deposited under the European Nucleotide Archive (PRJEB31487). We performed statistical analyses and graph plotting in R 3.5.2[7] using the dplyr[8], tidyr[9], ggplot2[10] and VennDiagram[11] packages.

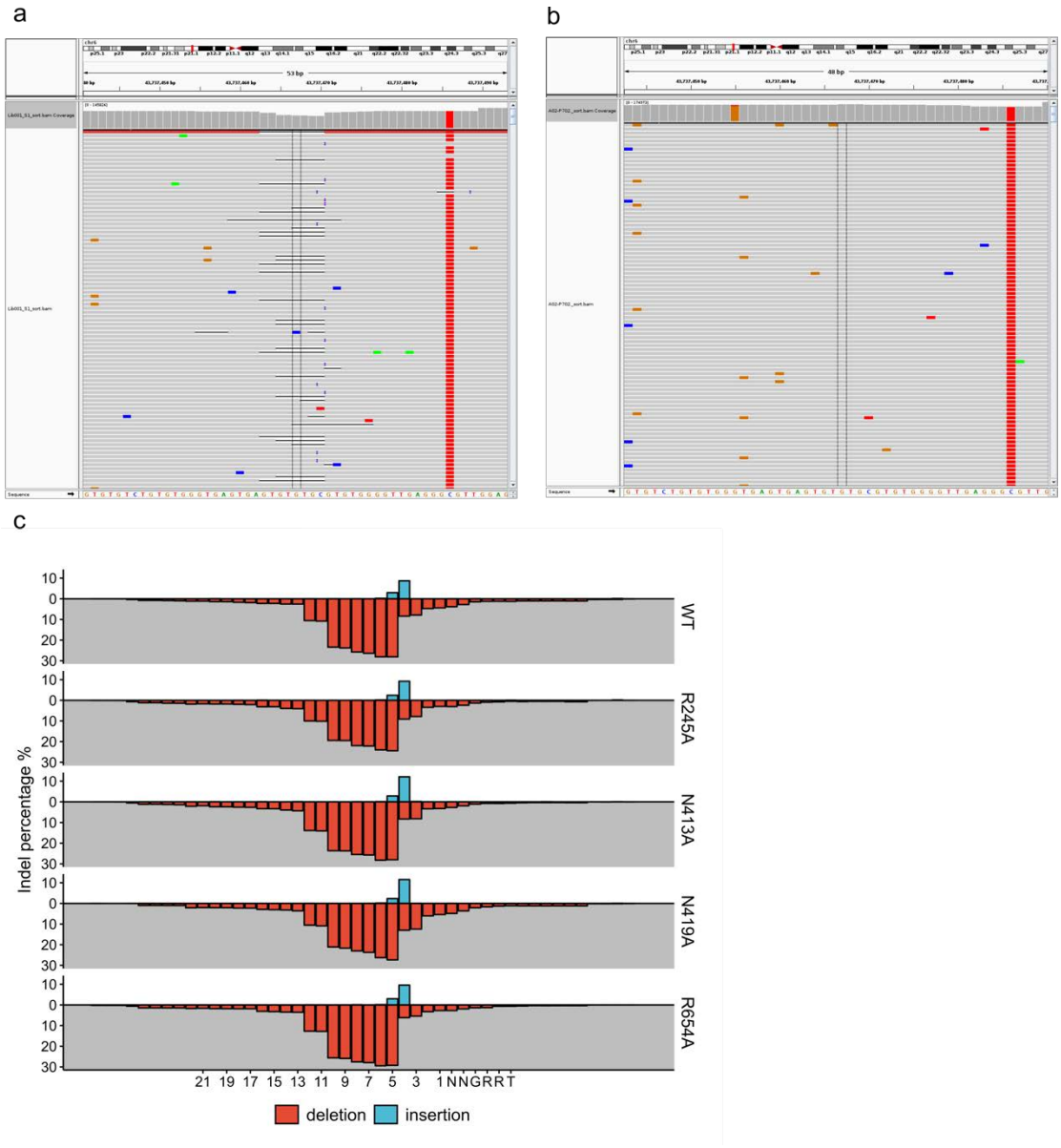

Fig. S1. Indel profile of SaCas9 targeting VEGFA_8 in HEK293T cells. IGV snapshots showing the insertions and deletions sequences of WT-SaCas9 editing in HEK293T (**a**) and the respective negative control of untreated cells (**b**). Indel profiles (**c**) showing per base insertion and deletion percentages (y-axis) introduced by WT-SaCas9 and the four single mutants: R245A, N413A, N419A, and R654A. The x-axis specifies the protospacer position (21-1) followed by the NNGRRT PAM site.

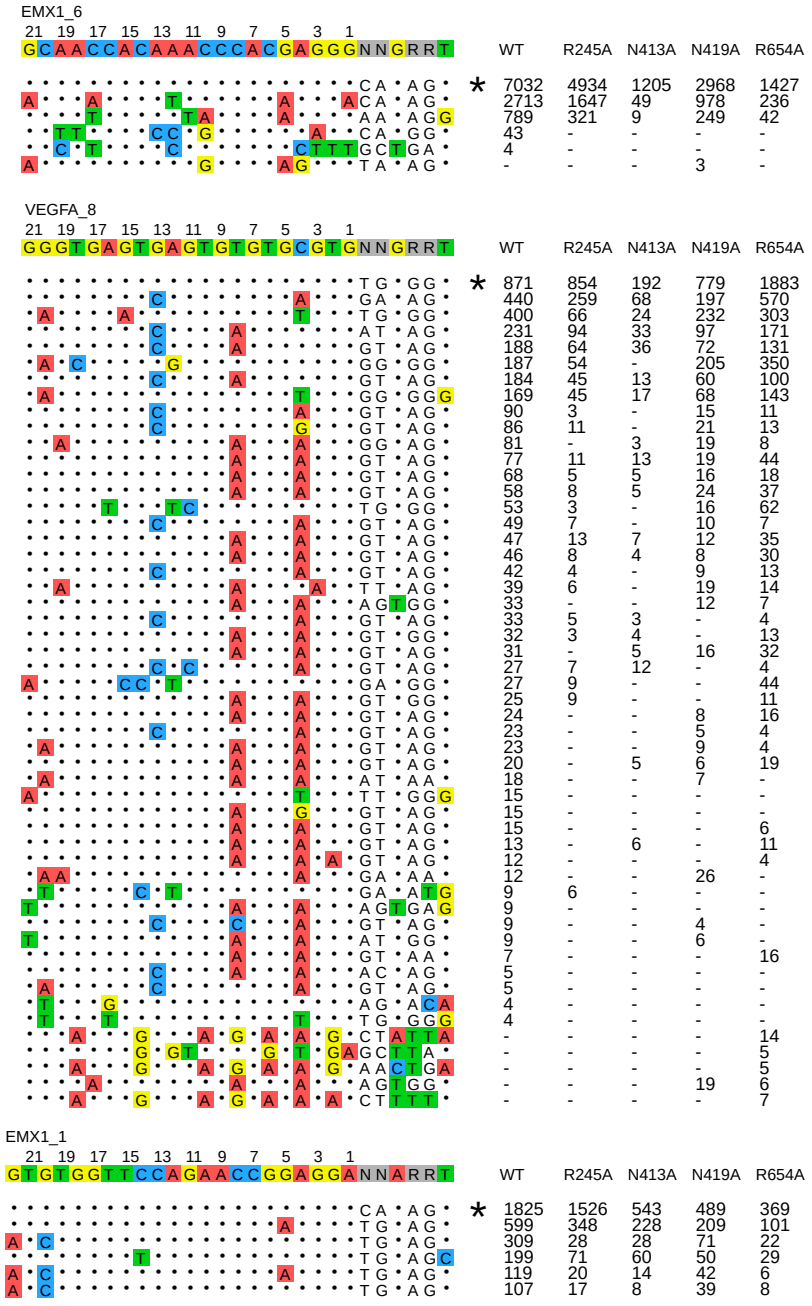

Fig. S2: Genome-wide editing results of WT and single-substitution SaCas9 variants at EMX1_6, VEGFA_8, and EMX1_1 using GUIDE-seq.
Genome-wide cleavage sites detected by GUIDE-seq. Read counts listed in the right represent number of GUIDE-seq reads. On target site is indicated with '*'. Mismatched bases in off-target sites with the on-target site are highlighted.

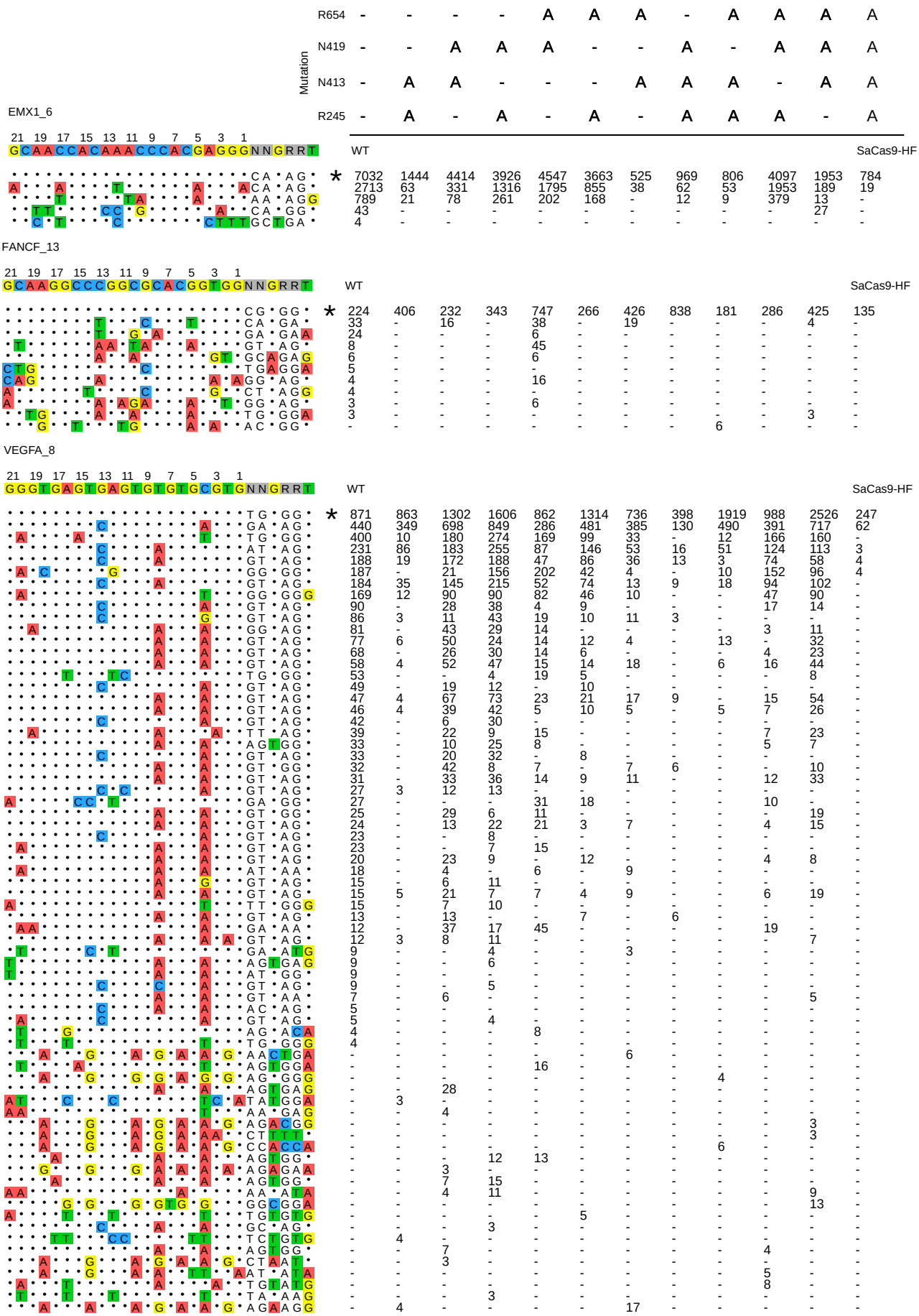

Fig. S3. Epistasis effect of SaCas9 residues on targeting specificity. Genome-wide cleavage of WT, double, triple, and quadruple mutant SaCas9s (mutation combination indicated in the top right panel) detected by GUIDE-seq at sites EMX1_6, FANCF_13 and VEGFA_8. Read counts listed in the right represent the cleavage frequency at a given site; on target site is marked with * for each sgRNA; mismatched positions are highlighted within the spacer or PAM.

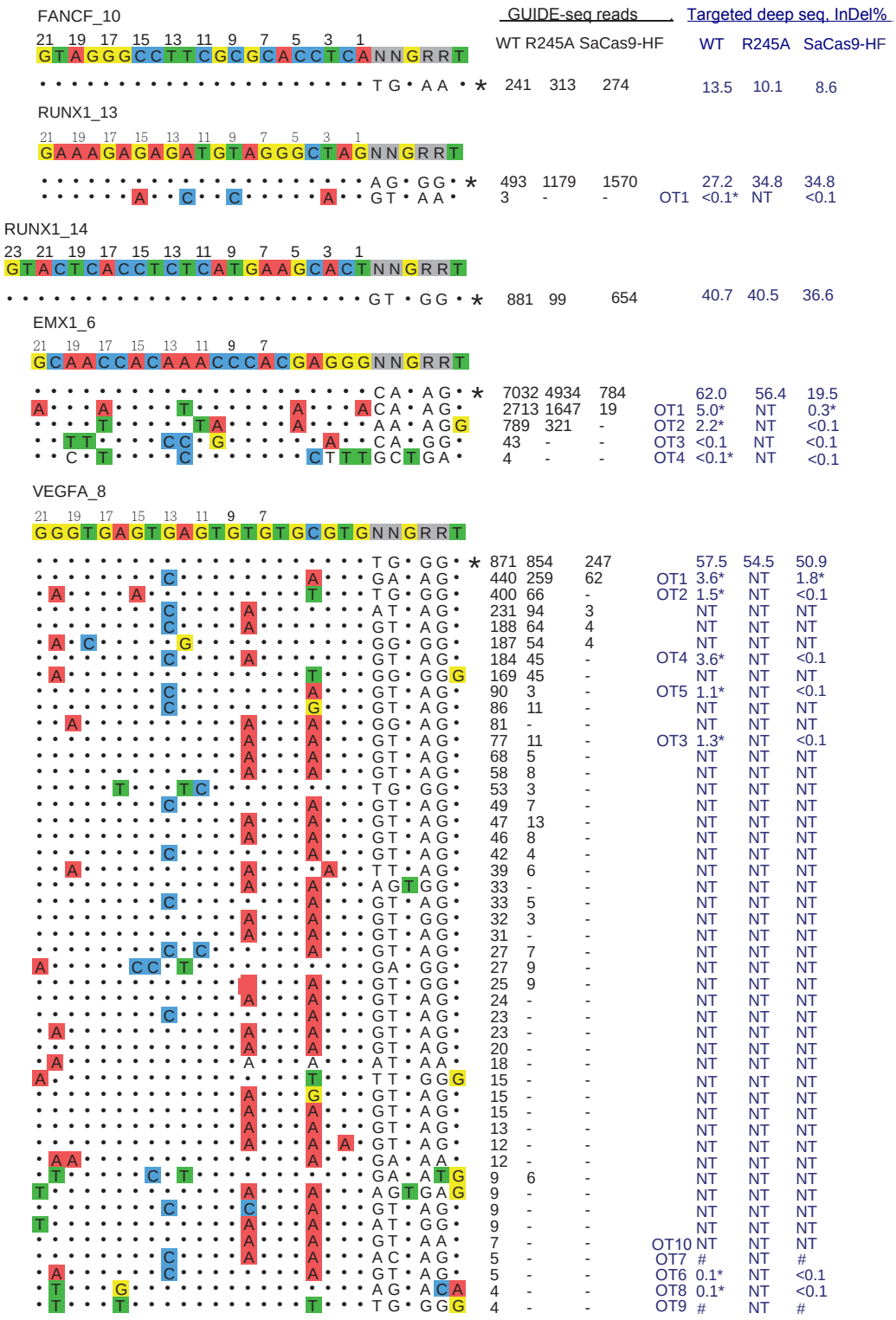

Fig. S4: Genome-wide editing specificity of wild-type SaCas9, R245A and quadruple mutants at five target sites with canonical NNGRRT PAM.

GUIDE-seq reads and InDel% detected in targeted deep sequencing for the on-target and off-target sites (in dark blue) are listed on the right. On target site is indicated with '*' at the right of the target sequence. Mismatched bases in off-target sites with the on-target site are highlighted. "NT" indicates off-target sites not subjected to targeted deep sequencing; "#" indicates off-target sites prone to false positives in targeted deep sequencing and percentages were not calculated; and InDel% marked with "*" indicates edited reads confirmed by IGV visualization.

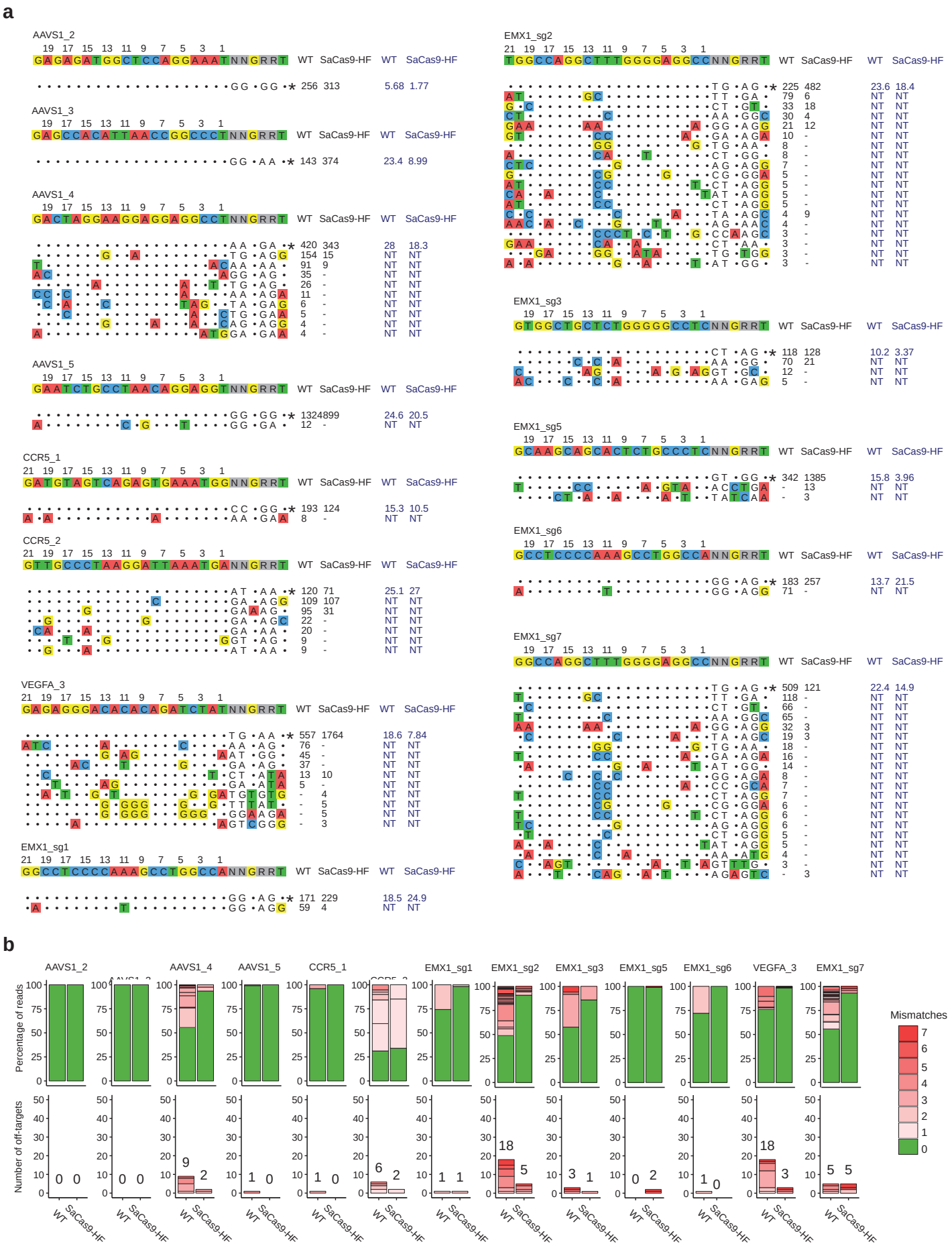

Fig. S5: **Genome-wide on- and off-target activities of wild-type SaCas9 and SaCas9-HF when targeting additional 13 human endogenous sites with canonical NNGRRT PAMs.**

a, Genome-wide cleavage sites detected by GUIDE-seq. On target site is indicated with '*' at the right of the target sequence. Mismatched bases in off-target sites with the on-target site are highlighted. InDel% detected in targeted deep sequencing for the on-target (in dark blue) are listed. **b**, Percentage of edited reads detected by GUIDE-seq at on-target site (green) and off-target sites (ordered by number of mismatches) among total edited reads (top) and summary of numbers of off-targets at 13 NNGRRT-PAM sites by WT-SaCas9 and SaCas9-HF (bottom).

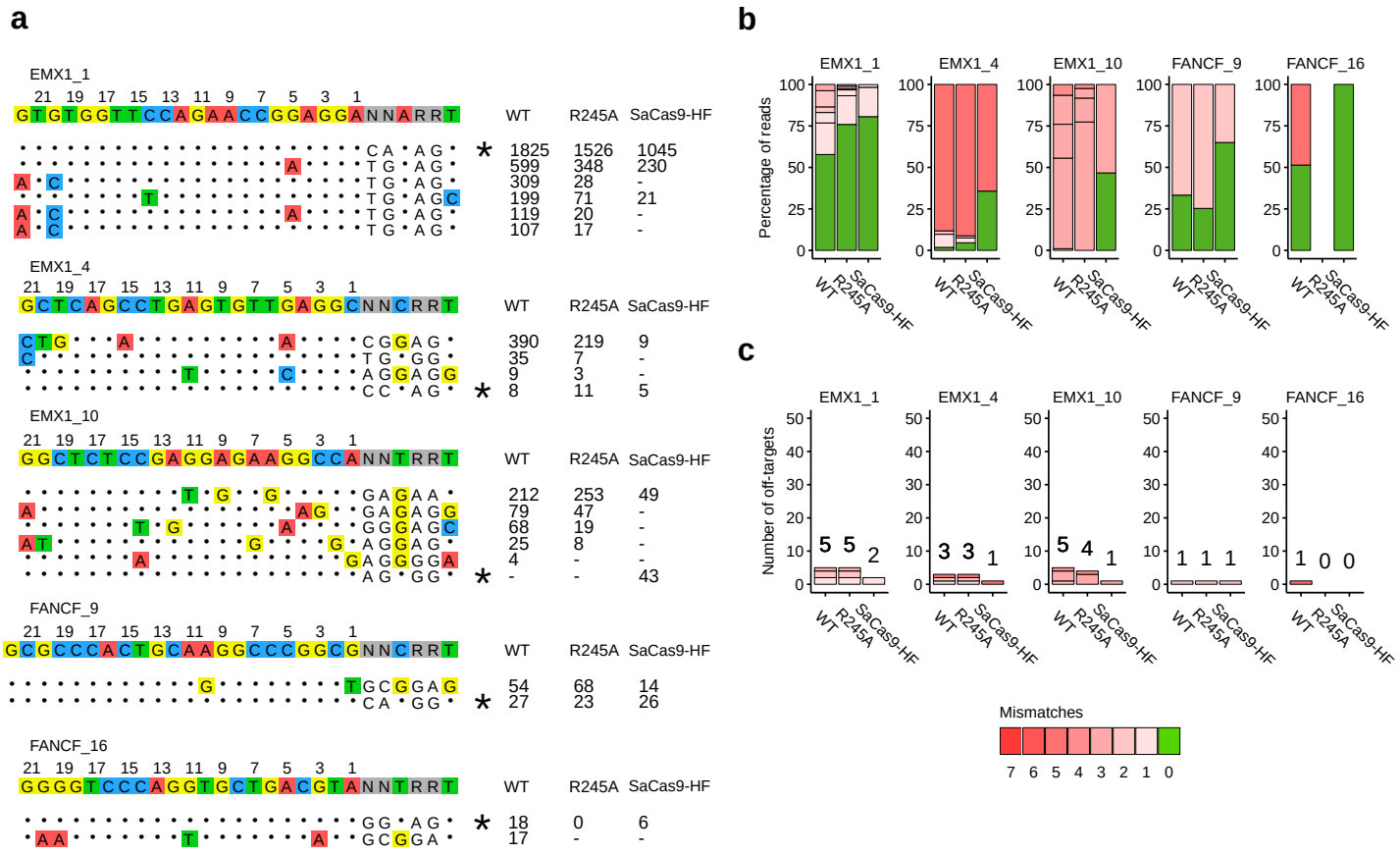

Fig. S6: **Genome-wide target profiles of WT-SaCas9, R245A and SaCas9-HF at non-canonical PAM sites.** (**a**) GUIDE-seq of the non-NNGRRT sites EMX1-1, EMX1_4, EMX1_10, FANCF_9 and FANCF_16. Read counts listed in the right represent the cleavage frequency at a given site; on target site is marked with * for each sgRNA; mismatched positions are highlighted within the spacer or PAM. (**b**) Summary bar chart of the genome-wide off-target detection of WT-SaCas9 (WT), SaCas9-R245A (R245A), SaCas9-HF (HF) at non-NNGRRT PAM sites: EMX1-1, EMX1_4, EMX1_10, FANCF_9 and FANCF_16. Each bar indicates the composition of dsODN integrated reads detected by GUIDE-seq for the SaCas9 variant; each stacked box shows the percentage of reads, as an approximation of cleavage frequency, at a target site with 0 (green) to 7 (red) mismatches. (**c**) A summary of the numbers of off-target sites.

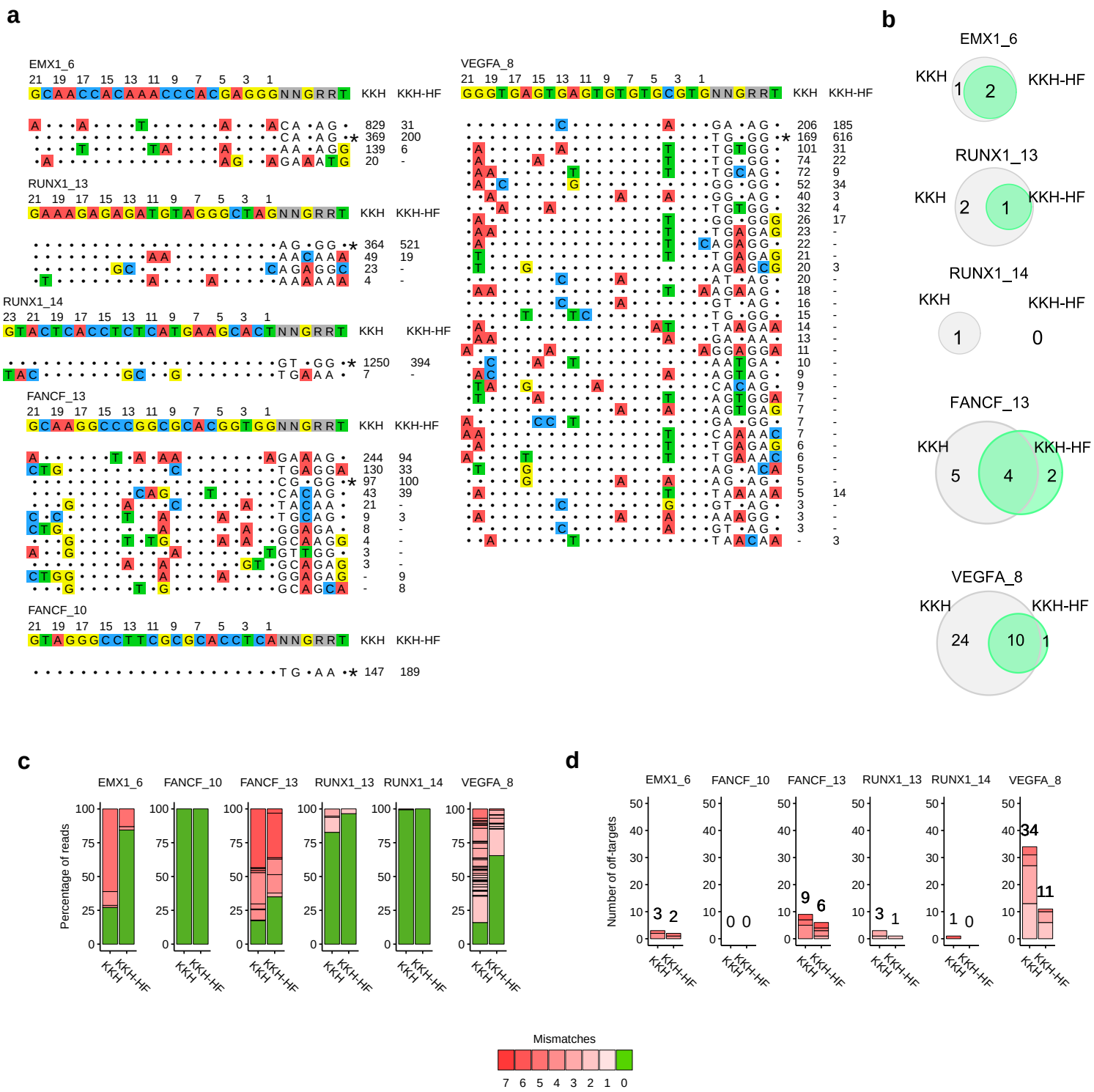

Fig. S7. On and off-target cleavage by KKH and KKH-HF at 6 canonical PAM human endogenous sites. **a**, GUIDE-seq detected cleavage sites by KKH-SaCas9 and KKH-HF. Read counts listed in the right represent number of GUIDE-seq reads. On-target site is indicated with '*'. Mismatched bases in off-target sites with the on-target site are highlighted. **b,** Venn diagram comparing the number of off-target sites between KKH-SaCas9 and KKH-HF in (a). **c**, Percentage of edited reads detected by GUIDE-seq at on-target site (green) and off-target sites (ordered by number of mismatches). **d**, Number of off-target sites in (c).

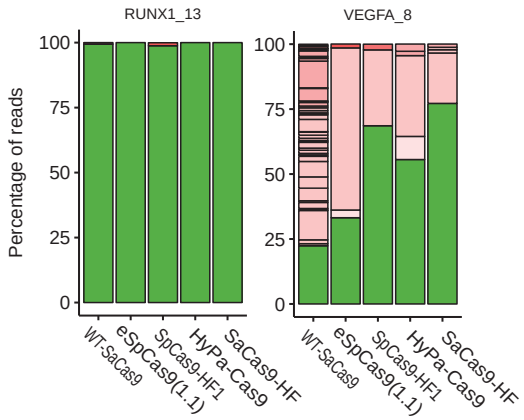

Fig. S8: Performance comparison of high fidelity Sa- and Sp-Cas9 variants.
Percentage of GUIDE-seq reads at on-target site (green) and off-target sites (ordered by number of mismatches) among total edited reads by each SaCas9 at each of the targeting site RUNX1_13 and VEGFA_8.

Table S1. Expression plasmids of SaCas9 variants used in this study.

Plasmid name	Mutation site
135-WT (BPK2139)	Wild-type SaCas9 without mutation.
135-2507-3011	R245A (2507 AG-GC), N413A (3011 AA-GC)
135-2507-3011-3029	R245A, N413A, N419A (3029 AA-GC)
135-2507-3011-3734	R245A, N413A, R654A (3734 AG-GC)
135-2507-3029	R245A, N419A
135-2507-3029-3734	R245A, N419A, R654A
135-2507-3734	R245A, R654A
135-3011-3029	N413A, N419A
135-3011-3029-3734	N413A, N419A, R654A
135-3011-3734	N413A, R654A
135-3029-3734	N419A, R654A
135-2507	R245A
135-3011	N413A
135-3029	N419A
135-3734	R654A
135-HF	R245A, N413A, N419A, R654A
135-S-HF	R499A, Q500A, R654A, G655A
135-KKH	G4118A, T4678A, G4818A
135-HF-KKH	R245A, N413A, N419A, R654A, G4118A, T4678A, G4818A

Table S2. Human endogenous sites targeted in this study.

Site	Spacer length (nt)	Genome location	Sequence of protospacer and PAM
EMX1_1	22	chr2+ 73160922-73160943	GTGTGGTTCCAGAACCGGAGGACAAAGT
EMX1_4	21	chr2+ 73160833-73160859	GCTCAGCCTGAGTGTTGAGGCCCCAGT
EMX1_6	21	chr2+ 73161089-73161109	GCAACCACAAACCCACGAGGGCAGAGT
EMX1_10	21	chr2- 73161274-73161300	GGCTCTCCGAGGAGAAGGCCAAGTGGT
FANCF_ 9	22	chr11+ 22647242-22647269	GCGCCCACTGCAAGGCCCGGCGCACGGT
FANCF_10	21	chr11- 22647350-22647376	GTAGGGCCTTCGCGCACCTCATGGAAT
FANCF _13	21	chr11+ 22647251-22647277	GCAAGGCCCGGCGCACGGTGGCGGGGT
FANCF_16	21	chr11+ 22647273-22647299	GGGGTCCCAGGTGCTGACGTAGGTAGT
RUNX1_13	21	chr21- 36421296-36421322	GAAAGAGAGATGTAGGGCTAGAGGGGT
RUNX1_14	23	chr21+ 36421131-36421159	GTACTCACCTCTCATGAAGCACTGTGGGT
VEGFA_8	21	chr6+ 43737453-43737473	GGGTGAGTGAGTGTGTGCGTGTGGGGT
AAVS1-2	20	chr19 + 55626992-55627017	GAGAGATGGCTCCAGGAAATGGGGGT
AAVS1-3	20	chr19 + 55627180-55627205	GAGCCACATTAACCGGCCCTGGGAAT
AAVS1-4	20	chr19+ 55627076-55627101	GACTAGGAAGGAGGAGGCCTAAGGAT
AAVS1-5	20	chr19+ 55627033-55627058	GAATCTGCCTAACAGGAGGTGGGGGT
CCR5-1	21	chr3 – 46413348-46413374	GATGTAGTCAGAGTGAAATGGCCGGGT
CCR5-2	21	chr3 + 46413579-46413605	GTTGCCCTAAGGATTAAATGAATGAAT
VEGFA 3	21	chr6- 43737522-43737548	GAGAGGGACACACAGATCTATTGGAAT
EMX1-sg1	21	chr2 – 73161167-73161193	GGCCTCCCCAAAGCCTGGCCAGGGAGT
EMX1-sg2	21	chr2 + 73161173-73161199	TGGCCAGGCTTTGGGGAGGCCTGGAGT
EMX1-sg3	20	chr2 + 73160858-73160883	GTGGCTGCTCTGGGGGCCTCCTGAGT
EMX1-sg5	20	chr2- 73161099-73161124	GCAAGCAGCACTCTGCCCTCGTGGGT
EMX1-sg6	20	chr2- 73161167-73161192	GCCTCCCCAAAGCCTGGCCAGGGAGT
EMX1-sg7	20	chr2+ 73161174-73161199	GGCCAGGCTTTGGGGAGGCCTGGAGT

+/- is the genomic orientation of sgRNA. PAM is highlighted in red.

Table S3. Target sequence in the spacer length and 5' mismatched G experiments

Site name	Length (bp)	Sequence
VEGFA-15-22	22	TGGGTGAGTGAGTGTGTGCGTG
VEGFA-15-21	21	GGGTGAGTGAGTGTGTGCGTG
VEGFA-15-20	20	GGTGAGTGAGTGTGTGCGTG
VEGFA-15-19	19	GTGAGTGAGTGTGTGCGTG
VEGFA-24-22	22	GGAGAGGGACACACAGATCTAT
VEGFA-24-21	21	GAGAGGGACACACAGATCTAT
VEGFA-24-20	20	AGAGGGACACACAGATCTAT
VEGFA-24-19	19	GAGGGACACACAGATCTAT
VEGFA-25-22	22	GCGTTGGAGCGGGGAGAAGGCC
VEGFA-25-21	21	CGTTGGAGCGGGGAGAAGGCC
VEGFA-25-20	20	GTTGGAGCGGGGAGAAGGCC
VEGFA-25-19	19	TTGGAGCGGGGAGAAGGCC
FANCF_13 (Mismatched G)		(G)GGCAAGGCCCGGCGCACGGTGGCGGGGT
RUNX1_13 (Mismatched G)		(G)GGAAAGAGAGATGTAGGGCTAGAGGGGT
FANCF_10 (Mismatched G)		(G)GGTAGGGCCTTCGCGCACCTCATGGAAT

Table S4. Primers for targeted deep sequencing.

Primer name	Sequence	
For PCR1		Note
VEGFA_8.fwd	TACACGACGCTCTTCCGATCTTGGGTGAATGGAGCGAGCAG	a
VEGFA_8.rev	TCCTCTCTATGGGCAGTCGGTGATGAGTGACCCCTGGCCTTCTC	a
EMX1_1.fwd	TACACGACGCTCTTCCGATCTGGCCTCCTGAGTTTCTCATCT	a
EMX1_1.rev	TCCTCTCTATGGGCAGTCGGTGATGACTCAGGCCCTTCCTCCT	a
EMX1_4.fwd	TACACGACGCTCTTCCGATCTAGCCTCAGTCTTCCCATCAG	a
EMX1_4.rev	TCCTCTCTATGGGCAGTCGGTGATTGGAACCACACCTTCACCTG	a
EMX1_6.fwd	TACACGACGCTCTTCCGATCTTCGATGTCACCTCCAATGAC	a
EMX1_6.rev	TCCTCTCTATGGGCAGTCGGTGATAGTGGCCAGAGTCCAGCTT	a
EMX1_10.fwd	TACACGACGCTCTTCCGATCTAGACACGGAGAGCAGCTG	a
EMX1_10.rev	TCCTCTCTATGGGCAGTCGGTGATCCATTGACAGAGGGACAAGC	a
FANCF_ 9.fwd	TACACGACGCTCTTCCGATCTCAAAGCGCCGATGGATGTG	a
FANCF_ 9.rev	TCCTCTCTATGGGCAGTCGGTGATGCGGTCTCAAGCACTACCTA	a
FANCF_10.fwd	TACACGACGCTCTTCCGATCTATCAGTACGCAGAGAGTCGC	a
FANCF_10.rev	TCCTCTCTATGGGCAGTCGGTGATACGTAGGTAGTGCTTGAGACC	a
FANCF _13.fwd	TACACGACGCTCTTCCGATCTCAAAGCGCCGATGGATGTG	a
FANCF _13.rev	TCCTCTCTATGGGCAGTCGGTGATGCGGTCTCAAGCACTACCTA	a
FANCF_16.fwd	TACACGACGCTCTTCCGATCTATGGATGTGGCGCAGGTAG	a
FANCF_16.rev	TCCTCTCTATGGGCAGTCGGTGATCATGGAATCCCTTCTGCAGC	a
RUNX1_13.fwd	TACACGACGCTCTTCCGATCTGAGTCCCAGAGGTATCCAGC	a
RUNX1_13.rev	TCCTCTCTATGGGCAGTCGGTGATTCCTCCTGAAAATGCACCCT	a
RUNX1_14.fwd	TACACGACGCTCTTCCGATCTCAAGCTGCCATTTCATTACAGG	a
RUNX1_14.rev	TCCTCTCTATGGGCAGTCGGTGATGAGGGTGCATTTTCAGGAGG	a
VEGFA_8.OT1.fwd	TACACGACGCTCTTCCGATCTGCTGTGCTATTGAGTGAAAAGT	b
VEGFA_8.OT1.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGCTTGCAACACTGAACGT	b
VEGFA_8.OT2.fwd	TACACGACGCTCTTCCGATCTGCTCATCCTGAAAGCAGAGG	b
VEGFA_8.OT2.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGATGGGCATTTATTGGAAA	b
VEGFA_8.OT3.fwd	TACACGACGCTCTTCCGATCTCAGCGTTTATATGATCTGGAGTAGA	b
VEGFA_8.OT3.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAACCATGGAGGTACAGTAAAACC	b
VEGFA_8.OT4.fwd	TACACGACGCTCTTCCGATCTTTTTAAAAAGTTTGTGTACCCTGA	b
VEGFA_8.OT4.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGCACTCACCATGAACAG	b
VEGFA_8.OT5.fwd	TACACGACGCTCTTCCGATCTGTGCTTGTGTACCTAAACGTATC	b
VEGFA_8.OT5.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATCCCCAAAAGTGCCTTCA	b
VEGFA_8.OT6.fwd	TACACGACGCTCTTCCGATCTGCCATCCTAGCCATCGTAAA	b
VEGFA_8.OT6.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTGGAGGGAACTTACCGTGA	b
VEGFA_8.OT7.fwd	TACACGACGCTCTTCCGATCTCAGAAAGCTGAGATTTATTATTGAAG	b
VEGFA_8.OT7.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTATAAAGTGGCCGACCTGGA	b
VEGFA_8.OT8.fed	TACACGACGCTCTTCCGATCTGCGTATTCAGGGTGTGCAAT	b
VEGFA_8.OT8.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTCTGGGAATCTAATGTATGGCA	b
VEGFA_8.OT9.fwd	TACACGACGCTCTTCCGATCTCCCGCCTCCAACTTCAAATG	b
VEGFA_8.OT9.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGCGCGTATGTATGTGTGT	b
EMX1_6.OT1.fwd	TACACGACGCTCTTCCGATCTTCCTCACACAAAAGCACATGT	b
EMX1_6.OT1.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGCCTATACATGCGACTGGT	b
EMX1_6.OT2.fwd	TACACGACGCTCTTCCGATCTACAAAACATAGTGCCTGCTCC	b
EMX1_6.OT2.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGAGCTTCTGTCTCTACCCC	b
EMX1_6.OT3.fwd	TACACGACGCTCTTCCGATCTCTCCCTCCTTCTTCCACAGG	b
EMX1_6.OT3.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGACAAACCCAACCTTTCCA	b
EMX1_6.OT4.fwd	TACACGACGCTCTTCCGATCTCATAGTGCACCTGTCCCTCA	b

EMX1_6.OT4.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGACTGAGGTGTTGTTGGC	b
FANCF_13.OT1.fwd	TACACGACGCTCTTCCGATCTGGTCACAAAAGATCTGGGGC	b
FANCF_13.OT1.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAAGCGCATGTCCACCTAAA	b
FANCF_13.OT2.fwd	TACACGACGCTCTTCCGATCTCAAACTCAAAAGGCTGCATCA	b
FANCF_13.OT2.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCATAGAGCCTCTCCTTCCCT	b
FANCF_13.OT3.fwd	TACACGACGCTCTTCCGATCTAACGAGAACTTTCATTTATGCCA	b
FANCF_13.OT3.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGACTAATTGACTACCCACTCACT	b
FANCF_13.OT4.fwd	TACACGACGCTCTTCCGATCTATTGTGGGGCAGGTACAGAG	b
FANCF_13.OT4.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGAATTGTGGTGTGTCTGCA	b
FANCF_13.OT5.fwd	TACACGACGCTCTTCCGATCTGAGCAGGATCTTCAGCACCT	b
FANCF_13.OT5.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGAGTACCTCCTGAGCCGC	b
FANCF_13.OT6.fwd	TACACGACGCTCTTCCGATCTCCCGGCTCCCAGGTTCGCTGA	b
FANCF_13.OT6.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCGTCCGGACCCTCCAAGTG	b
FANCF_13.OT7.fwd	TACACGACGCTCTTCCGATCTACCCTAGTGCTGTGACCAC	b
FANCF_13.OT7.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGGTACAGAGTCCAGGCAG	b
FANCF_13.OT8.fwd	TACACGACGCTCTTCCGATCTCGTGGAAAAGTGGCACCC	b
FANCF_13.OT8.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGCCAACCCCAGGAATATTG	b
FANCF_13.OT9.fwd	TACACGACGCTCTTCCGATCTAAGCACAACCAAGGAGAGGG	b
FANCF_13.OT9.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTCTTCAGTCACCTCACCAGT	b
FANCF_13.OT10.fwd	TACACGACGCTCTTCCGATCTTGGGGTTTCATGCTCTCTTCA	b
FANCF_13.OT10.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGTCCACCCTTTCATCCTCC	b
RUNX1_13.OT1.fwd	TACACGACGCTCTTCCGATCTACTGGCTTGGATTCCTTCTCA	b
RUNX1_13.OT1.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGGGCGTCAGATTGAAACC	b
VEGFA_15.fwd	TACACGACGCTCTTCCGATCTTGGGTGAATGGAGCGAGCAG	b
VEGFA_15.rev	TCCTCTCTATGGGCAGTCGGTGATGAGTGACCCCTGGCCTTCTC	b
VEGFA_24.fwd	TACACGACGCTCTTCCGATCTGAGCCGTTCCCTCTTTGCTA	b
VEGFA_24.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGAGTGAGGACGTGTGTGT	b
VEGFA_25.fwd	TACACGACGCTCTTCCGATCTGCAGCGTCTTCGAGAGTGAGG	b
VEGFA_25.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGGGGAGAGGGACACACAGAT	b
AAVS1_2.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCTGGCTCCATCGTAAGCAAACCTTA	b
AAVS1_2.fwd	TACACGACGCTCTTCCGATCTGGTCTAACCCCCACCTCCTGTTA	b
AAVS1_3.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCACTGTGGGGTGGAGGGGACA	b
AAVS1_3.fwd	TACACGACGCTCTTCCGATCTGGATCCTGTGTCCCCGAGCTG	b
AAVS1_4.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACAGGAGGTGGGGGTTAGACC	b
AAVS1_4.fwd	TACACGACGCTCTTCCGATCTGCCACTAGGGACAGGATTGGTGACAG	b
AAVS1_5.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCAGGAAATGGGGGTGTGTCACC	b
AAVS1_5.fwd	TACACGACGCTCTTCCGATCTAGGCCTCCTCCTTCCTAGTCTC	b
CCR5_1.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCATCTGTGGTGGCAGACGAAACATTT	b
CCR5_1.fwd	TACACGACGCTCTTCCGATCTAGGCAAGACCATCAGATGTTTGGTGA	b
CCR5_2.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTGTGCTTCAAGGTCCTTGTCTGCAA	b
CCR5_2.fwd	TACACGACGCTCTTCCGATCTACATGCACTATGAGCAAGCCAGTAAT	b
EMX1_sg1.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGTGGGCCCAAGCTGGACTC	b
EMX1_sg1.fwd	TACACGACGCTCTTCCGATCTATTGCTTGTCCCTCTGTCAATGGC	b
EMX1_sg2.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGTGGGCCCAAGCTGGACTCT	b
EMX1_sg2.fwd	TACACGACGCTCTTCCGATCTGCCTCAGCCAGCCCATTGC	b
EMX1_sg3.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCAGTCTTCCCATCAGGCTCTCAGCT	b
EMX1_sg3.fwd	TACACGACGCTCTTCCGATCTGGGAGGGAGGGGCACAGATGA	b
EMX1_sg5.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGCCAATGGGGAGGACATCGATGT	b
EMX1_sg5.fwd	TACACGACGCTCTTCCGATCTTCCAGCTTGGGCCCACGCA	b
EMX1_sg6.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGTGGGCCCAAGCTGGACTC	b

EMX1_sg6.fwd	TACACGACGCTCTTCCGATCTATTGCTTGTCCCTCTGTCAATGGC	b
EMX1_sg7.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGTGGGCCCAAGCTGGACTC	b
EMX1_sg7.fwd	TACACGACGCTCTTCCGATCTATTGCTTGTCCCTCTGTCAATGGC	b
VEGFA_3.rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGAGAAGGCCAGGGGTCAC	b
VEGFA_3.fwd	TACACGACGCTCTTCCGATCTGGCAGGGGAAGGCGGAGAG	b
For PCR2		
MI.P501	AATGATACGGCGACCACCGAGATCTACACTAGATCGCNNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T	
MI.P502	AATGATACGGCGACCACCGAGATCTACACCTCTCTATNNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T	
MI.P503	AATGATACGGCGACCACCGAGATCTACACTATCCTCTNNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T	
MI.P504	AATGATACGGCGACCACCGAGATCTACACAGAGTAGANNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T	
MI.P505	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGNNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T	
MI.P506	AATGATACGGCGACCACCGAGATCTACACACTGCATANNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T	
MI.P507	AATGATACGGCGACCACCGAGATCTACACAAGGAGTANNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T	
MI.P508	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTNNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T	
P7-I-1	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTGACTGGAGTTCAGACGTGT	
P7-I-2	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTGACTGGAGTTCAGACGTGT	
P7-I-3	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTGACTGGAGTTCAGACGTGT	
P7-I-4	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTGACTGGAGTTCAGACGTGT	
P7-I-5	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTGACTGGAGTTCAGACGTGT	
P7-I-6	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTGACTGGAGTTCAGACGTGT	
P7-I-7	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTGACTGGAGTTCAGACGTGT	
P7-I-8	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTGACTGGAGTTCAGACGTGT	
P7-I-9	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTGACTGGAGTTCAGACGTGT	
P7-I-10	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTGACTGGAGTTCAGACGTGT	
P7-I-11	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTGACTGGAGTTCAGACGTGT	
P7-I-12	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTGACTGGAGTTCAGACGTGT	
P7-I-13	CAAGCAGAAGACGGCATACGAGATAACTTCACGTGACTGGAGTTCAGACGTGT	
P7-I-14	CAAGCAGAAGACGGCATACGAGATTGGAGAGGGTGACTGGAGTTCAGACGTGT	
P7-I-15	CAAGCAGAAGACGGCATACGAGATACGCATCGGTGACTGGAGTTCAGACGTGT	
P7-I-16	CAAGCAGAAGACGGCATACGAGATGTACCGTTGTGACTGGAGTTCAGACGTGT	
P7-I-17	CAAGCAGAAGACGGCATACGAGATTACAGTTAGTGACTGGAGTTCAGACGTGT	
P7-I-18	CAAGCAGAAGACGGCATACGAGATAATCAACTGTGACTGGAGTTCAGACGTGT	
P7-I-19	CAAGCAGAAGACGGCATACGAGATGTACCTAGGTGACTGGAGTTCAGACGTGT	
P7-I-20	CAAGCAGAAGACGGCATACGAGATCTGGAACAGTGACTGGAGTTCAGACGTGT	
P7-I-21	CAAGCAGAAGACGGCATACGAGATGGTGACTAGTGACTGGAGTTCAGACGTGT	
P7-I-22	CAAGCAGAAGACGGCATACGAGATGTGCAACCGTGACTGGAGTTCAGACGTGT	
P7-I-23	CAAGCAGAAGACGGCATACGAGATGCCTGTCTGTGACTGGAGTTCAGACGTGT	
P7-I-24	CAAGCAGAAGACGGCATACGAGATACTGATGGGTGACTGGAGTTCAGACGTGT	
P7-I-25	CAAGCAGAAGACGGCATACGAGATATGCTAACGTGACTGGAGTTCAGACGTGT	
P7-I-26	CAAGCAGAAGACGGCATACGAGATCACTGAGTGTGACTGGAGTTCAGACGTGT	
P7-I-27	CAAGCAGAAGACGGCATACGAGATTAGGCCATGTGACTGGAGTTCAGACGTGT	
P7-I-28	CAAGCAGAAGACGGCATACGAGATCAGCAGTCGTGACTGGAGTTCAGACGTGT	
P7-I-29	CAAGCAGAAGACGGCATACGAGATTTCTGAGAGTGACTGGAGTTCAGACGTGT	
P7-I-30	CAAGCAGAAGACGGCATACGAGATGGACGTTAGTGACTGGAGTTCAGACGTGT	

P7-I-31	CAAGCAGAAGACGGCATACGAGATGTGTAGGTGTGACTGGAGTTCAGACGTGT	
P7-I-32	CAAGCAGAAGACGGCATACGAGATCATCTCAGGTGACTGGAGTTCAGACGTGT	
P7-I-33	CAAGCAGAAGACGGCATACGAGATGCATAGCAGTGACTGGAGTTCAGACGTGT	
P7-I-34	CAAGCAGAAGACGGCATACGAGATCAGTGCACGTGACTGGAGTTCAGACGTGT	
P7-I-35	CAAGCAGAAGACGGCATACGAGATTTCGGCATGTGACTGGAGTTCAGACGTGT	
P7-I-36	CAAGCAGAAGACGGCATACGAGATCAACAGGTGTGACTGGAGTTCAGACGTGT	
P7-I-37	CAAGCAGAAGACGGCATACGAGATAACACTCGGTGACTGGAGTTCAGACGTGT	
P7-I-38	CAAGCAGAAGACGGCATACGAGATGTCCTGACGTGACTGGAGTTCAGACGTGT	
P7-I-39	CAAGCAGAAGACGGCATACGAGATGACGTAGAGTGACTGGAGTTCAGACGTGT	
P7-I-40	CAAGCAGAAGACGGCATACGAGATGATTGGCAGTGACTGGAGTTCAGACGTGT	
P7-I-41	CAAGCAGAAGACGGCATACGAGATGCCACGACGTGACTGGAGTTCAGACGTGT	
P7-I-42	CAAGCAGAAGACGGCATACGAGATTTGTTACGGTGACTGGAGTTCAGACGTGT	
P7-I-43	CAAGCAGAAGACGGCATACGAGATACGACCTAGTGACTGGAGTTCAGACGTGT	
P7-I-44	CAAGCAGAAGACGGCATACGAGATTGATAATGGTGACTGGAGTTCAGACGTGT	
P7-I-45	CAAGCAGAAGACGGCATACGAGATGGTTCCATGTGACTGGAGTTCAGACGTGT	
P7-I-46	CAAGCAGAAGACGGCATACGAGATCCAGTATCGTGACTGGAGTTCAGACGTGT	
P7-I-47	CAAGCAGAAGACGGCATACGAGATGTCCAGCTGTGACTGGAGTTCAGACGTGT	
P7-I-48	CAAGCAGAAGACGGCATACGAGATTAACCTTCGTGACTGGAGTTCAGACGTGT	
P7-I-49	CAAGCAGAAGACGGCATACGAGATATAGGCTGGTGACTGGAGTTCAGACGTGT	
P7-I-50	CAAGCAGAAGACGGCATACGAGATCCACTTGAGTGACTGGAGTTCAGACGTGT	
P7-I-51	CAAGCAGAAGACGGCATACGAGATTGGACCACGTGACTGGAGTTCAGACGTGT	
P7-I-52	CAAGCAGAAGACGGCATACGAGATTGCGTCAGGTGACTGGAGTTCAGACGTGT	
P7-I-53	CAAGCAGAAGACGGCATACGAGATGTCGTGTAGTGACTGGAGTTCAGACGTGT	
P7-I-54	CAAGCAGAAGACGGCATACGAGATGGAGGCCAGTGACTGGAGTTCAGACGTGT	
P7-I-55	CAAGCAGAAGACGGCATACGAGATCTAGCGTGGTGACTGGAGTTCAGACGTGT	
P7-I-56	CAAGCAGAAGACGGCATACGAGATGCGCGTGTGTGACTGGAGTTCAGACGTGT	
P7-I-57	CAAGCAGAAGACGGCATACGAGATCATTATGGGTGACTGGAGTTCAGACGTGT	
P7-I-58	CAAGCAGAAGACGGCATACGAGATATTATACGGTGACTGGAGTTCAGACGTGT	
P7-I-59	CAAGCAGAAGACGGCATACGAGATTTAGACTCGTGACTGGAGTTCAGACGTGT	
P7-I-60	CAAGCAGAAGACGGCATACGAGATGTACGCACGTGACTGGAGTTCAGACGTGT	
P7-I-61	CAAGCAGAAGACGGCATACGAGATGAGCCATCGTGACTGGAGTTCAGACGTGT	
P7-I-62	CAAGCAGAAGACGGCATACGAGATGTGGAGTCGTGACTGGAGTTCAGACGTGT	
P7-I-63	CAAGCAGAAGACGGCATACGAGATCAGCTTCGGTGACTGGAGTTCAGACGTGT	
P7-I-64	CAAGCAGAAGACGGCATACGAGATCACAAGTAGTGACTGGAGTTCAGACGTGT	
P7-I-65	CAAGCAGAAGACGGCATACGAGATGATGTGGTGTGACTGGAGTTCAGACGTGT	
P7-I-66	CAAGCAGAAGACGGCATACGAGATCTTGGAGGGTGACTGGAGTTCAGACGTGT	
P7-I-67	CAAGCAGAAGACGGCATACGAGATCGAACGTCGTGACTGGAGTTCAGACGTGT	
P7-I-68	CAAGCAGAAGACGGCATACGAGATATTGTCAAGTGACTGGAGTTCAGACGTGT	
P7-I-69	CAAGCAGAAGACGGCATACGAGATGTTCCGGAGTGACTGGAGTTCAGACGTGT	
P7-I-70	CAAGCAGAAGACGGCATACGAGATCCTAACGGGTGACTGGAGTTCAGACGTGT	
P7-I-71	CAAGCAGAAGACGGCATACGAGATATCTGGACGTGACTGGAGTTCAGACGTGT	
P7-I-72	CAAGCAGAAGACGGCATACGAGATCTATAGACGTGACTGGAGTTCAGACGTGT	
P7-I-73	CAAGCAGAAGACGGCATACGAGATCTTGCGGCGTGACTGGAGTTCAGACGTGT	

P7-I-74	CAAGCAGAAGACGGCATACGAGATAATTAGGCGTGACTGGAGTTCAGACGTGT	
P7-I-75	CAAGCAGAAGACGGCATACGAGATCGTGTAGTGTGACTGGAGTTCAGACGTGT	
P7-I-76	CAAGCAGAAGACGGCATACGAGATTGCCAACAGTGACTGGAGTTCAGACGTGT	
P7-I-77	CAAGCAGAAGACGGCATACGAGATTACACAAGGTGACTGGAGTTCAGACGTGT	
P7-I-78	CAAGCAGAAGACGGCATACGAGATTAACTAGGGTGACTGGAGTTCAGACGTGT	
P7-I-79	CAAGCAGAAGACGGCATACGAGATATAGTGTTGTGACTGGAGTTCAGACGTGT	
P7-I-80	CAAGCAGAAGACGGCATACGAGATATGTAATAGTGACTGGAGTTCAGACGTGT	
P7-I-81	CAAGCAGAAGACGGCATACGAGATACTTGGTGGTGACTGGAGTTCAGACGTGT	
P7-I-82	CAAGCAGAAGACGGCATACGAGATCAGTCTAGGTGACTGGAGTTCAGACGTGT	
P7-I-83	CAAGCAGAAGACGGCATACGAGATTGACGCCTGTGACTGGAGTTCAGACGTGT	
P7-I-84	CAAGCAGAAGACGGCATACGAGATACCATGTAGTGACTGGAGTTCAGACGTGT	
P7-I-85	CAAGCAGAAGACGGCATACGAGATTTACCGACGTGACTGGAGTTCAGACGTGT	
P7-I-86	CAAGCAGAAGACGGCATACGAGATCCAGAAGTGTGACTGGAGTTCAGACGTGT	
P7-I-87	CAAGCAGAAGACGGCATACGAGATAGGTTGCTGTGACTGGAGTTCAGACGTGT	
P7-I-88	CAAGCAGAAGACGGCATACGAGATCTGAGTGAGTGACTGGAGTTCAGACGTGT	
P7-I-89	CAAGCAGAAGACGGCATACGAGATGGTACTGTGTGACTGGAGTTCAGACGTGT	
P7-I-90	CAAGCAGAAGACGGCATACGAGATTTGGTGCCGTGACTGGAGTTCAGACGTGT	
P7-I-91	CAAGCAGAAGACGGCATACGAGATTTCGGATGGTGACTGGAGTTCAGACGTGT	
P7-I-92	CAAGCAGAAGACGGCATACGAGATCCGTATATGTGACTGGAGTTCAGACGTGT	
P7-I-93	CAAGCAGAAGACGGCATACGAGATTATCGACAGTGACTGGAGTTCAGACGTGT	
P7-I-94	CAAGCAGAAGACGGCATACGAGATATAGCCGAGTGACTGGAGTTCAGACGTGT	
P7-I-95	CAAGCAGAAGACGGCATACGAGATGATCGGTTGTGACTGGAGTTCAGACGTGT	
P7-I-96	CAAGCAGAAGACGGCATACGAGATAATGGTTCGTGACTGGAGTTCAGACGTGT	

a. PCR2: use one of the MIP.5##s and one of the P7##s (list in Table S5). Requires custom sequencing primers Index 1 and Read2 (list in Table S5).

b. PCR2: use one of the MIP.5##s and one of the P7-I-##s in PCR2. Compatible with standard **I**llumina sequencing primers.

* Indicates a phosphorothioate bond modification.

Table S5. Adaptor and primer sequences for GUIDE-seq.

Note: Each adaptor is formed by using one TOP and one corresponding BOT oligos, with the following annealing program: 95°C for 1 min; slow ramping down at -0.2°C/sec to 4°C.

TOP.ID	TOP.oligo.seq	BOT.ID	BOT.oligo.seq
TOP001	TACACGACGCTCTTCCGATCTNNWNNWNATGCCANTGCCGTT	BOT001	/5Phos/ACGGCANTGGCAT/3Phos/
TOP002	TACACGACGCTCTTCCGATCTNNWNNWNACCATCNCAACGAT	BOT002	/5Phos/TCGTTGNGATGGT/3Phos/
TOP003	TACACGACGCTCTTCCGATCTNNWNNWNGTGGCCNTACTACT	BOT003	/5Phos/GTAGTANGGCCAC/3Phos/
TOP004	TACACGACGCTCTTCCGATCTNNWNNWNGGAGTTNGAGGTGT	BOT004	/5Phos/CACCTCNAACTCC/3Phos/
TOP005	TACACGACGCTCTTCCGATCTNNWNNWNATTGCANAGCAACT	BOT005	/5Phos/GTTGCTNTGCAAT/3Phos/
TOP006	TACACGACGCTCTTCCGATCTNNWNNWNTACAACNCGAGTAT	BOT006	/5Phos/TACTCGNGTTGTA/3Phos/
TOP007	TACACGACGCTCTTCCGATCTNNWNNWNTGCGTTNCTAGCGT	BOT007	/5Phos/CGCTAGNAACGCA/3Phos/
TOP008	TACACGACGCTCTTCCGATCTNNWNNWNTGTTCCNTCTCACT	BOT008	/5Phos/GTGAGANGGAACA/3Phos/
TOP009	TACACGACGCTCTTCCGATCTNNWNNWNTGCAGTNCCTCGAT	BOT009	/5Phos/TCGAGGNACTGCA/3Phos/
TOP010	TACACGACGCTCTTCCGATCTNNWNNWNCGATACNACTGCCT	BOT010	/5Phos/GGCAGTNGTATCG/3Phos/
TOP011	TACACGACGCTCTTCCGATCTNNWNNWNTCTGCGNAGTCTGT	BOT011	/5Phos/CAGACTNCGCAGA/3Phos/
TOP012	TACACGACGCTCTTCCGATCTNNWNNWNTCTTGCNGGAGTCT	BOT012	/5Phos/GACTCCNGCAAGA/3Phos/
TOP013	TACACGACGCTCTTCCGATCTNNWNNWNAGGCTTNACGTGTT	BOT013	/5Phos/ACACGTNAAGCCT/3Phos/
TOP014	TACACGACGCTCTTCCGATCTNNWNNWNTCACGANGTCACAT	BOT014	/5Phos/TGTGACNTCGTGA/3Phos/
TOP015	TACACGACGCTCTTCCGATCTNNWNNWNAGCCGGNAGAGTAT	BOT015	/5Phos/TACTCTNCCGGCT/3Phos/
TOP016	TACACGACGCTCTTCCGATCTNNWNNWNATGGAANGGTGGCT	BOT016	/5Phos/GCCACCNTTCCAT/3Phos/
TOP017	TACACGACGCTCTTCCGATCTNNWNNWNAGACACNCAATGTT	BOT017	/5Phos/ACATTGNGTGTCT/3Phos/
TOP018	TACACGACGCTCTTCCGATCTNNWNNWNGGTAGTNTCATAGT	BOT018	/5Phos/CTATGANACTACC/3Phos/
TOP019	TACACGACGCTCTTCCGATCTNNWNNWNATCGAGNGATCTAT	BOT019	/5Phos/TAGATCNCTCGAT/3Phos/
TOP020	TACACGACGCTCTTCCGATCTNNWNNWNCAACAANTGCCAAT	BOT020	/5Phos/TTGGCANTTGTTG/3Phos/
TOP021	TACACGACGCTCTTCCGATCTNNWNNWNGTGCAANCCAATCT	BOT021	/5Phos/GATTGGNTTGCAC/3Phos/
TOP022	TACACGACGCTCTTCCGATCTNNWNNWNACAATGNTCACAGT	BOT022	/5Phos/CTGTGANCATTGT/3Phos/
TOP023	TACACGACGCTCTTCCGATCTNNWNNWNGGCTCTNAACGTAT	BOT023	/5Phos/TACGTTNAGAGCC/3Phos/
TOP024	TACACGACGCTCTTCCGATCTNNWNNWNCTCCACNATTCCTT	BOT024	/5Phos/AGGAATNGTGGAG/3Phos/
TOP025	TACACGACGCTCTTCCGATCTNNWNNWNAAGGCGNCTCCTTT	BOT025	/5Phos/AAGGAGNCGCCTT/3Phos/
TOP026	TACACGACGCTCTTCCGATCTNNWNNWNGAGACANGTGGAAT	BOT026	/5Phos/TTCCACNTGTCTC/3Phos/
TOP027	TACACGACGCTCTTCCGATCTNNWNNWNATGCGTNAATGCAT	BOT027	/5Phos/TGCATTNACGCAT/3Phos/
TOP028	TACACGACGCTCTTCCGATCTNNWNNWNTACAAGNTGGTCCT	BOT028	/5Phos/GGACCANCTTGTA/3Phos/
TOP029	TACACGACGCTCTTCCGATCTNNWNNWNACTCCTNTGTGTTT	BOT029	/5Phos/AACACANAGGAGT/3Phos/
TOP030	TACACGACGCTCTTCCGATCTNNWNNWNTGGAGCNCTTGTCT	BOT030	/5Phos/GACAAGNGCTCCA/3Phos/
TOP031	TACACGACGCTCTTCCGATCTNNWNNWNCAACTGNTCAGACT	BOT031	/5Phos/GTCTGANCAGTTG/3Phos/
TOP032	TACACGACGCTCTTCCGATCTNNWNNWNTCAGATNACCAGCT	BOT032	/5Phos/GCTGGTNATCTGA/3Phos/
TOP033	TACACGACGCTCTTCCGATCTNNWNNWNTGGCCGNTTACTGT	BOT033	/5Phos/CAGTAANCGGCCA/3Phos/
TOP034	TACACGACGCTCTTCCGATCTNNWNNWNGGTGCTNAATCACT	BOT034	/5Phos/GTGATTNAGCACC/3Phos/
TOP035	TACACGACGCTCTTCCGATCTNNWNNWNCATACGNATACAGT	BOT035	/5Phos/CTGTATNCGTATG/3Phos/
TOP036	TACACGACGCTCTTCCGATCTNNWNNWNTGAATANCCTGGCT	BOT036	/5Phos/GCCAGGNTATTCA/3Phos/
TOP037	TACACGACGCTCTTCCGATCTNNWNNWNGTGGTCNATCGTAT	BOT037	/5Phos/TACGATNGACCAC/3Phos/
TOP038	TACACGACGCTCTTCCGATCTNNWNNWNGAACCTNATGACAT	BOT038	/5Phos/TGTCATNAGGTTC/3Phos/
TOP039	TACACGACGCTCTTCCGATCTNNWNNWNACACGANCTATAGT	BOT039	/5Phos/CTATAGNTCGTGT/3Phos/
TOP040	TACACGACGCTCTTCCGATCTNNWNNWNATATGCNGAGACTT	BOT040	/5Phos/AGTCTCNGCATAT/3Phos/
TOP041	TACACGACGCTCTTCCGATCTNNWNNWNCGCTTANGTGCTGT	BOT041	/5Phos/CAGCACNTAAGCG/3Phos/
TOP042	TACACGACGCTCTTCCGATCTNNWNNWNACTACTNGAGGATT	BOT042	/5Phos/ATCCTCNAGTAGT/3Phos/
TOP043	TACACGACGCTCTTCCGATCTNNWNNWNGCTCCGNACCATAT	BOT043	/5Phos/TATGGTNCGGAGC/3Phos/
TOP044	TACACGACGCTCTTCCGATCTNNWNNWNTTCGGCNATAGTGT	BOT044	/5Phos/CACTATNGCCGAA/3Phos/
TOP045	TACACGACGCTCTTCCGATCTNNWNNWNTTAGAGNCCATGCT	BOT045	/5Phos/GCATGGNCTCTAA/3Phos/
TOP046	TACACGACGCTCTTCCGATCTNNWNNWNAGGTGANGTTCTAT	BOT046	/5Phos/TAGAACNTCACCT/3Phos/
TOP047	TACACGACGCTCTTCCGATCTNNWNNWNAACATTNGCAGGTT	BOT047	/5Phos/ACCTGCNAATGTT/3Phos/
TOP048	TACACGACGCTCTTCCGATCTNNWNNWNGGTGGCNATGGAAT	BOT048	/5Phos/TTCCATNGCCACC/3Phos/

Primers used in enrichment PCR1 and PCR2

Primer	sequence	Note

OFF_GSP1_-	GGATCTCGACGCTCTCCCTGTTTAATTGAGTTGTCATATGTTAATAAC	for PCR1
OFF_GSP1_+	GGATCTCGACGCTCTCCCTATACCGTTATTAACATATGACA	for PCR1
OFF_GSP2_-	CCTCTCTATGGGCAGTCGGTGATACATATGACAACTCAATTAAAC	for PCR2
OFF_GSP2_+	CCTCTCTATGGGCAGTCGGTGATTTGAGTTGTCATATGTTAATAACGGTA	for PCR2
P558	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT	for PCR2
P701	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	for PCR2
P702	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P703	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P704	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P705	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P706	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P707	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P708	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P709	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P710	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P711	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P712	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P713	CAAGCAGAAGACGGCATACGAGATAACTTCACGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P714	CAAGCAGAAGACGGCATACGAGATTGGAGAGGGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P715	CAAGCAGAAGACGGCATACGAGATACGCATCGGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P716	CAAGCAGAAGACGGCATACGAGATGTACCGTTGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P717	CAAGCAGAAGACGGCATACGAGATTACAGTTAGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P718	CAAGCAGAAGACGGCATACGAGATAATCAACTGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P719	CAAGCAGAAGACGGCATACGAGATGTACCTAGGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P720	CAAGCAGAAGACGGCATACGAGATCTGGAACAGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P721	CAAGCAGAAGACGGCATACGAGATGGTGACTAGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P722	CAAGCAGAAGACGGCATACGAGATGTGCAACCGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P723	CAAGCAGAAGACGGCATACGAGATGCCTGTCTGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P724	CAAGCAGAAGACGGCATACGAGATACTGATGGGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P725	CAAGCAGAAGACGGCATACGAGATATGCTAACGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P726	CAAGCAGAAGACGGCATACGAGATCACTGAGTGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P727	CAAGCAGAAGACGGCATACGAGATTAGGCCATGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P728	CAAGCAGAAGACGGCATACGAGATCAGCAGTCGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P729	CAAGCAGAAGACGGCATACGAGATTTCTGAGAGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P730	CAAGCAGAAGACGGCATACGAGATGGACGTTAGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P731	CAAGCAGAAGACGGCATACGAGATGTGTAGGTGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P732	CAAGCAGAAGACGGCATACGAGATCATCTCAGGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P733	CAAGCAGAAGACGGCATACGAGATGCATAGCAGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P734	CAAGCAGAAGACGGCATACGAGATCAGTGCACGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P735	CAAGCAGAAGACGGCATACGAGATTTCGGCATGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P736	CAAGCAGAAGACGGCATACGAGATCAACAGGTGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P737	CAAGCAGAAGACGGCATACGAGATAACACTCGGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P738	CAAGCAGAAGACGGCATACGAGATGTCCTGACGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P739	CAAGCAGAAGACGGCATACGAGATGACGTAGAGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P740	CAAGCAGAAGACGGCATACGAGATGATTGGCAGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P741	CAAGCAGAAGACGGCATACGAGATGCCACGACGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P742	CAAGCAGAAGACGGCATACGAGATTTGTTACGGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P743	CAAGCAGAAGACGGCATACGAGATACGACCTAGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P744	CAAGCAGAAGACGGCATACGAGATTGATAATGGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P745	CAAGCAGAAGACGGCATACGAGATGGTTCCATGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P746	CAAGCAGAAGACGGCATACGAGATCCAGTATCGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P747	CAAGCAGAAGACGGCATACGAGATGTCCAGCTGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
P748	CAAGCAGAAGACGGCATACGAGATTAACCTTCGTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	
Index1	ATCACCGACTGCCCATAGAGAGGACTCCAGTCAC (Custom sequencing primer Index1)	
Read2	GTGACTGGAGTCCTCTCTATGGGCAGTCGGTGAT (Custom sequencing primer Read2)	

Reference

1. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–95 (2010).
2. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–198 (2015).
3. Zheng, Z. *et al.* Anchored multiplex PCR for targeted next-generation sequencing. *Nat. Med.* **20**, 1479–1484 (2014).
4. Bushnell, B. BBMap (version 37.75) [Software]. *Available at https://sourceforge.net/projects/bbmap/* (2015).
5. Tsai, S. Q., Topkar, V. V, Joung, J. K. & Aryee, M. J. Open-source guideseq software for analysis of GUIDE-seq data. *Nat. Biotechnol.* (2016). doi:10.1038/nbt.3534
6. Minoche, A. E., Dohm, J. C. & Himmelbauer, H. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. *Genome Biol.* **12**, R112 (2011).
7. Team, R. C. R: A Language and Environment for Statistical Computing. *Vienna, Austria* (2018).
8. Wickham, H. & Francois, R. Dplyr:Dplyr: a Grammar of Data Manipulation. *Cran* (2014).
9. Wickham, H. & Henry, L. tidyr: Easily Tidy Data with ‘spread()’ and ‘gather()’ Functions. *R Packag. version 0.8.0. https//CRAN.R-project.org/package=tidyr* (2018).
10. Valero-Mora, P. M. ggplot2: Elegant Graphics for Data Analysis . *J. Stat. Softw.* (2015). doi:10.18637/jss.v035.b01
11. Chen, H. & Boutros, P. C. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* **12**, 35 (2011).